Fusion of Vision, GPS and 3D Gyro Data in Solving Camera Registration Problem for Direct Visual Navigation

Zhencheng Hu[*1] Keiichi Uchimura[*2]
Computer Science Dept., Kumamoto Univ. [*1]
(2-39-1 Kurokami, Kumamoto, (096)342-3894, E-mail:hu@cs.kumamoto-u.ac.jp)
Dept. of Systems and Information, Graduate School of Science and Technology, Kumamoto Univ. [*2]
(2-39-1 Kurokami, Kumamoto, (096)342-3638, E-mail:uchimura@cs.kumamoto-u.ac.jp)

This paper presents a precise and robust camera registration solution for the novel vision-based road navigation system - VICNAS, which superimposes virtual 3D navigation indicators and traffic signs upon the real road view in an Augmented Reality (AR) space. Traditional vision based or inertial sensor based solutions of registration problem are mostly designed for well-structured environment, which is however unavailable in a wide-open uncontrolled road environment for navigation purposes. This paper proposed a hybrid system that combines computer vision, GPS and 3D inertial gyroscope technologies to provide precise and robust camera pose estimation. The fusion approach is based on our PMM (parameterized model matching) algorithm, in which the road shape model is derived from the digital map data, and matched with road features extracted from real images. Inertial data estimates the initial possible motion, and also serves as relative tolerance to stable the pose output. The algorithms proposed in this paper are validated with the experimental results of real road tests under different road conditions.

Keywords: *Data fusion, camera registration problem, augmented reality, vehicle navigation system.*

1. Introduction

Tracking a moving camera's three-dimensional (3D) position and orientation is essential to the so-called registration problem in an Augmented Reality Context. The objects in the real and virtual world must be properly aligned with respect to each other, which requires knowing the observer's exact 3D viewing pose (position and orientation) data. Especially when the observer (camera) is moving, accurate estimation of the 3D pose data and tracking the temporal coherence from successive images will absolutely affect the synthesizing accuracy and visual performance of virtual objects in the AR space.

To deal with this problem, many approaches have been proposed in recent years[1,2,3]. Previous work in this area can be divided into three main categories: 1) solutions based on external tracking devices like inertial sensors, beacons or transponders, 2) image processing solutions that directly estimate camera pose from the same imagery observed by the viewer, 3) hybrid solutions attempt to overcome the drawbacks of any single sensing solution.

Inertial sensors are widely used for motion tracking[4,5]. With the characteristics of self-contained, source-less and high sampling rate, they are suitable for tracking the rapid motions like vehicle or aviation movement. However, since inertial sensors only measure the variation rate or accelerations, the output signals have to be integrated to obtain the position and orientation data. As a result, longer integrated time produces significant accumulated drift because of noise or bias.

The general concept of vision-based camera 3D pose estimation is to find the best set of camera position and orientation data (the six extrinsic parameters) to fit a known model in the target image[6,7,8]. Unlike other sensing technologies, vision solutions directly estimate camera pose from the same imagery that is also used as the real world background, therefore vision solutions always offer the best visual perceived performance when the virtual objects are projected to the background. However, vision solutions also suffer from the high computational cost, sensitive to noise and lack of robustness since they depend on image feature extraction and tracking result.

Hybrid solutions are widely applied in recent research works since different sensors can be used to compensate others limitation. Chai et al.[9] employs an adaptive pose estimator with vision and inertial sensors for overcoming the problems of inertial sensor drift and vision sensor slow measurement. The extended Kalman filter (EKF) is used for data fusion and error compensation. You et al. [10] also combined vision and inertial sensor with a two-channel complementary EKF, which can take advantage of the low-frequency stability of vision sensors and the high-frequency tracking of gyro sensors. However, most of these approaches are designed for well-structured environment. Especially for the vision sensors, predefined artificial markers are vital for feature tracking process, which is however unavailable in the outdoor uncontrolled road navigation environment.

For on-road navigation applications, the fast translation movement along vehicle's moving direction results in the continuous change of image background. In addition, there are no predefined squares or circle

markers that can be constantly tracked in the wide-opened real road scene. Territory map, some landmarks and road lane-markers are the only features that can be used to determine camera's pose. Because of these factors, this registration problem cannot be solved by previous hybrid algorithms.

In this paper, we extended our previous work[3] of pure vision based solution to a hybrid solution that combines vision, GPS and 3D inertial gyroscope sensing technologies. We still restrict our aim at the registration problem for on-road navigation applications. The fusion approach is based on our PMM (parameterized model matching) algorithm, in which the road shape model is derived from the digital map referring to GPS absolute road position, and matches with road features extracted from the real image. Inertial data estimates the initial state of searching parameters, and also serves as relative tolerance to stable the pose output. Comparing with the previous hybrid algorithms, the proposed solution employs GPS and inertial sensor to obtain absolute position, which leads to the proper road position by Map-matching process. Additionally, PMM algorithm matches road lane markers with the road shape model derived from road map, which makes our algorithm very robust to the featureless road environment.

The paper is organized as follows: the new concept of vision-based road navigation system is quickly reviewed in Section 2. Section 3 describes our hybrid data fusion solution that combines computer vision, GPS and 3D gyro data to solve camera registration problem. Section 4 gives the implementation details of superimposing virtual navigation indicators and traffic signs upon real road view based on the estimation result from Section 3. Experimental results of real road and discussions are described in Section 5.

2. Review of New Road Navigation Concept

With the development of voice guidance and dynamical traffic information exchange techniques, recent vehicle navigation systems will guide you with voice instructions well in advance of your next move along a pre-planned route. However even with the voice guidance and digital road map, a driver still has to compare by himself the road scene ahead with his digital map to determine which lane to take or, at which intersection to turn. It is not only inconvenient, but also even dangerous in some cases, especially during the high-speed driving in dense traffic roads. A new concept of direct visual navigation and its prototype system – Vision-based Car Navigation System (VICNAS) was proposed by the authors[3] to overcome this inconvenient problem of current navigation system. As shown in Fig. 1, VICNAS employs Augmented Reality technique to superimpose virtual direction indicators and traffic information bulletins upon the real driver's view.

Since all the virtual indicators and overlay graphics have to be aligned properly with the real road scene from driver's view, the accuracy of navigation that VICNAS can provide absolutely depends on the accuracy of the estimated viewing pose, which means camera registration accuracy directly determines the visually-perceived performance of AR system.

There are several factors that have to be considered to solve the Registration problem for VICNAS. Road navigation is mainly used for high speed moving vehicles, which gives a fast translation along vehicle's moving axis. No predefined square or circle markers can be put in the wide-open real road scene. Territory map, some landmarks and road shapes are the only features that can be used to determine camera pose. Even small drift in camera pose will lead a significant displacement of virtual objects on the projected image.

Fig. 1. Prototype driver interface of VICNAS system

3. Hybrid Solution of Camera Registration

Our previous work[3] employed pure vision-based solution for camera registration. It shared similar goals with the video-based model tracking solution described by Valinetti[11]. Valinetti introduced a scalar evaluation score based on the local image gradient along the projected model lines to evaluate the existence possibility of certain camera pose values. In our Parameterized Model Matching (PMM) algorithm, we chose road shape as the target model since it can be directly derived from the digital road map and is fairly easy to track in different lighting conditions. It simplified the 2D-3D feature corresponding problem to a 2D-2D model matching optimization and showed good visual perceived performance.

However as described in Section 1, like other pure vision-based solutions, it suffered from the lack of robustness since it depends on image features extraction and tracking result. To overcome the problem, this paper proposes a hybrid data fusion solution that combines vision, GPS and 3D inertial gyroscope sensing technologies to provide precise and robust camera pose estimation. Fig. 2 shows the basis block diagram of the solution. Absolute road position is derived from the fusion of GPS and gyro data. Road Modeling Block

(RMB) uses digital road map data to generate a shape model of roads ahead from this position. It will match with the road features extracted from real image and output the estimation result. The angular rate data obtained from the gyro sensor initializes the possible motion, and also serves as relative tolerance to stable the final output.

3.1 Absolute Road Positioning

In an open, well-communicated environment, accuracy of differential GPS (DGPS) sensor can achieve 1.5m horizontally and 5m in altitude. In urban area, high buildings and signal random reflection (so-called multi-path) will significantly affect GPS accuracy. In this case, we use inertial sensors to compensate the GPS data. Most navigations systems will use map-matching algorithm to pull the absolute positioning data to the nearest possible road according to moving trace history.

Since GPS sampling rate (1Hz~10Hz) normally is lower than the inertial sensor sampling rate (10Hz~500Hz), the fusion of GPS and 3D gyro for absolute road position is based on a predictor-corrector control theory as shown in Fig. 3.

GPS data and gyro data are fed into evaluation module and integration module separately. After checking data integrity, captured satellites number and DOP value, every evaluated trustable GPS data will start a new loop and reset gyro's integrating module. The difference between new GPS position and integrated gyro's predication will be fed back into the gyro integration module as a dynamical correction factor.

Assuming $P_g(t_i) = (X_{t_i}, Y_{t_i}, Z_{t_i})$ are evaluated trustable GPS position data, where $t_i = t_0, t_1, ..., t_n$ and $V_i(\tau) = (vx_\tau, vy_\tau, vz_\tau)$ are velocity data integrated from 3D gyro's acceleration output. Then the absolute road positioning output between two trustable GPS data $P_g(t_n)$ and $P_g(t_{n+1})$ can be calculated by:

$$P(t) = P_g(t_n) + \left[\int_{t_n}^{t} V_i(\tau) d\tau + \Delta P_{t_n}(t - t_n) \right] \quad (1)$$

where ΔP_{t_n} is the feedback adjustment factor to correct 3D gyro data.

$$\Delta P_{t_n} = \frac{1}{t_n - t_{n-1}} \left\{ P_g(t_n) - \left[\int_{t_{n-1}}^{t_n} V_i(\tau) d\tau + \Delta P_{t_{n-1}}(t_n - t_{n-1}) \right] \right\} \quad (2)$$

3.2 Reference Frames for Vision System

There are five coordinate systems involved in VICNAS[3]: World Coordinate System (WCS), Vehicle Coordinate System (VCS), Camera Coordinate System (CCS), Inertial Coordinate System (GCS) and Projected Image Coordinate System (ICS). As shown in Fig. 4, the origin of WCS is located on road centerline as the current Map-matching result by comparing digital road map with GPS/gyroscope output. Y-axis of WCS is on the tangent direction of road centerline, Z-axis points at up and X-axis points at left. Assuming local road surface is flat, VCS can be treated as the relative WCS with an offset and heading angle on the ground. More generally, the relative position of WCS and VCS can be described as a rigid motion of translation and rotation.

Mapping from CCS to ICS is a perspective projection, while transformation from VCS to CCS can also be described as a rigid motion of translation and rotation. Therefore, the homogeneous transformation matrix between WCS and ICS is shown in eq.(3).

$$\kappa \vec{p}_i = \Gamma \vec{P}_C = K[E \mid 0] \vec{P}_C = K[E \mid 0] \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \vec{P}_W \quad (3)$$

where $\vec{p}_i, \vec{P}_C$ and $\vec{P}_W$ are the homogeneous coordinates of ICS, VCS and WCS respectively. κ is an arbitrary scale factor, K is called camera intrinsic parameters matrix, and R , T are the rotation matrix and translation vector respectively.

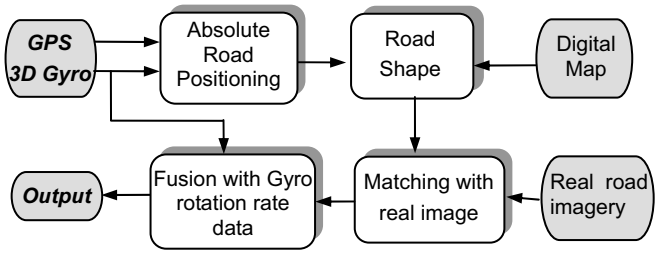

Fig. 2. Block diagram of our hybrid solution

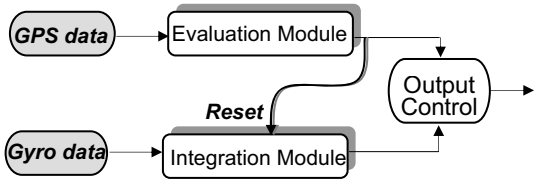

Fig. 3. Fusion of GPS and 3D gyro for road positioning

Fixed camera intrinsic parameters can be easily obtained from the initial calibration. Therefore the camera pose estimation problem is to compute the 6 extrinsic parameters in essence. If we collect the 6 camera extrinsic parameters in one vector σ , we can simply parameterize the perspective mapping relationship between the 2D image coordinates in ICS and the 3D world coordinates in WCS as follows:

$$\vec{p}_i = \mathbf{\Gamma}(\vec{P}_W; \boldsymbol{\sigma}). \quad (4)$$

In general, each matching pairs of $(\vec{p}_i, \vec{P}_W)$ will contribute to the determination of camera pose vector σ . Since it is impossible to obtain the 3D information

directly from image data along, we adopted parameterized model matching algorithm to transfer the direct matching problem to an optimized problem of searching the best camera pose set.

3.3 Road Shape Model

The information of roads ahead from the current road position obtained in Section 3.1 can be extracted from the 2D digital navigation map. Road skeleton node positions and the associated attributes (road name, construction level, direction information and lanes number in either direction) are employed to build the road shape model.

We used clothoid curves to fit the road shape ahead[12]. Eq. (5) is a very compact parameterized multi-lane model on WCS:

$$\Re_i = (c_{0i}, c_{1i}, n_{li}, n_{ri}, w_i, L_i)^T \tag{5}$$

where n_{li} and n_{ri} are numbers of road lanes on each side, w_i is the average lane width during this segment. c_{0i}, c_{1i} and L_i are clothoid shape parameters. Since this road model's origin is based on the road central skeleton line, we have to transfer it to the vehicle coordinate system according to the current road position's offset and heading angle. This model will then be projected to the 2D driver's view by the perspective mapping from eq.(4).

$$\vec{p}(\Re_i) = \boldsymbol{\Gamma}(\Re_i; \boldsymbol{\sigma}) \tag{6}$$

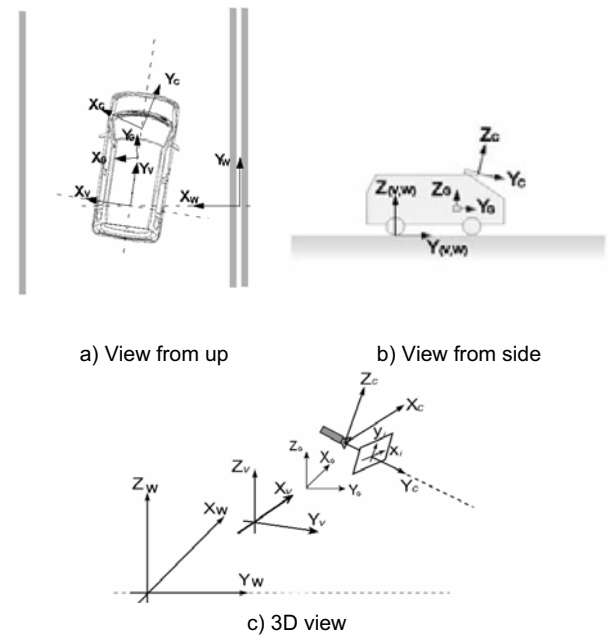

Fig. 4. Reference frames for vision system

3.4 Model Matching

Varying camera pose vector σ will generate different projected road shapes according to eq.(6), thus the camera pose estimation problem has been transferred to the optimized problem of searching a best camera pose vector σ to match with road image data. Road lane markers are extracted from the road images by the WLE (White Line Emphasis) filter developed by Oike[16]. WLE filter compares the clockwise moment and local summation result of the traditional differential filter [-1, 1], and output the enhanced lane markers. The filter kernel size is Nx1, where N corresponds to lane markers maximum width in the image. Pavement boundaries are also extracted as adjutant road shape information[17].

Gray scale correlation is not preferred in the matching process due to various types and colors of road lane markers, different lighting and weather conditions as well. To counteract this effect, a Road Shape Look-up table (RSL) is employed to give peak values at the position of lane marker, and lower values at their neighbors[14].

Therefore, a normalized camera pose estimation score function can be given as:

$$\begin{aligned} E(\boldsymbol{\sigma}) &= \frac{1}{|\eta_{\boldsymbol{\sigma}}|} \sum_{p \in \eta_\sigma} \left\| RSL :\rightarrow p(x,y) \right\| \\ &= \frac{1}{|\eta_{\boldsymbol{\sigma}}|} \sum_{p \in \eta_\sigma} \left\| RSL :\rightarrow \boldsymbol{\Gamma}(\Re; \boldsymbol{\sigma}) \right\| \end{aligned} \tag{7}$$

where η_σ represents the set of points belonging to the perspective road model $\Re$. Every point $p(x,y)$ on the road model that has a non-zero RSL value will contribute to the score function. In other word, the maximum of estimation score will be reached at the perfect matching of projective road model to the road shapes on the image.

Theoretically, the whole region that is lower than the disappearing line will be the candidate region for searching the lane markers. However in practice, we only scan the regions centered by previous extraction results of lane markers because of the constraint of road shape continuity. With the general camera and lens setup, the farthest road lane marker we can detect (width >1 pixel) will be 80 ~ 100 meters.

A direct search algorithm is adopted in the optimization searching operation[13], while the fusion of vision and gyro data gives out its initial state and searching range. Since the gyro data is defined in the inertial coordinate system, it is necessary to convert it to the world coordinate system. Let $\Omega(\theta, \phi, \psi)$ be the absolute rotation angle (Euler angle), and $W(\omega_x, \omega_y, \omega_z)$ represent the angular rate from inertial

sensor output. According to [15], the conversion from inertial angular rate to the world will be

$$\Delta\Omega = \begin{bmatrix} \dot{\theta} \\ \dot{\phi} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ \tan\theta\sin\phi & 1 & -\tan\theta\cos\phi \\ -\sin\phi/\cos\theta & 0 & \cos\phi/\cos\theta \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} \quad (8)$$

Eq. (8) is employed to predict the motion of image features. Estimated inertial angular data will be served as the initial state and relative searching tolerance in the Direct Search algorithm for vision based model-matching process.

Gyroscope data will be directly adopted as sensor fusion output if one of the following conditions could be satisfied. 1) No optimized parameter vector can be found in the searching range. 2) There is less or no white lane markers can be extracted from input image. 3) Variation of image based estimated result has exceeded gyro's accuracy tolerance.

4. Experiment Result and Discussions

A SONY analog CCD camera was mounted on the front roof of test vehicle. Image sequences were recorded in NTSC format at the frame rate of 30fps. Differential GPS data (Trimble® AgGPS) and inertial data (DataTech® GU-3023) were sent to PC's serial port and recorded at the frequency of 10Hz and 60Hz seperately. Zenrin® Z-Map was used as the 2D road map.

As the first phase of VICNAS project, our tests were based on the off-line processing. Road tests were carried out on different kinds of road (express toll-way, city highway, downtown street and countryside road), different lane structures (one-way or two-way, 1~6 lanes, with or without central separators) and shapes (straight, curve, S-curve). We assume that white lane markers have been painted in most part of the test road. Bad weather conditions like snow, heavy rain and fog are not considered in this test. Sample images are shown in Fig.5.

4.1 Data Fusion Result of GPS and Gyro

As described in section 3.1, high buildings and signal random reflection in urban area will significantly affect GPS accuracy. A typical DGPS data error is shown in Fig. 6, where "■" points are DGPS data and a significant discontinuous jump can be found on the left bottom side. "●" points are data fusion result where gyro data is adopted to compensate GPS data. The evaluation module ejects most of the unreliable GPS points and interpolates the output with gyro data. In this particular example, all GPS data with less than 7 captured satellites and which DOP value is higher than 2.1 will be ejected.

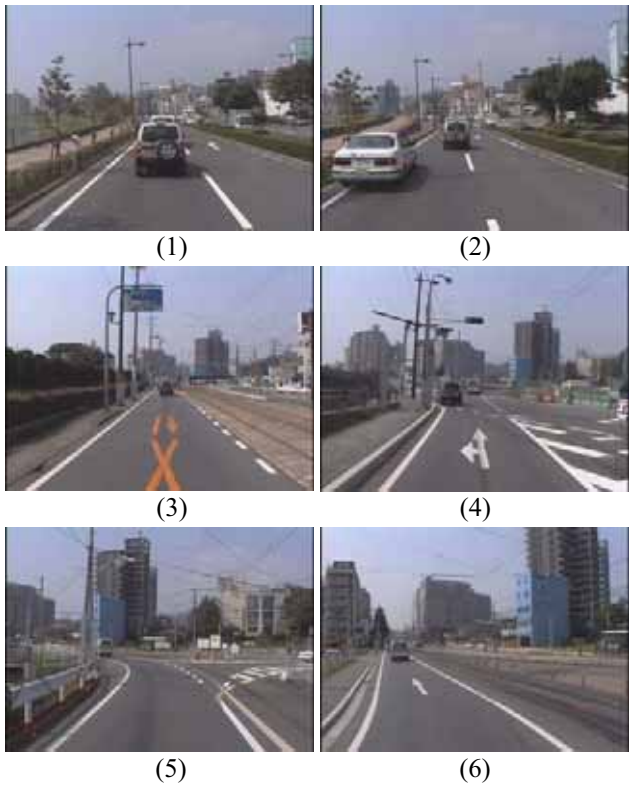

Fig. 5. Test road environment

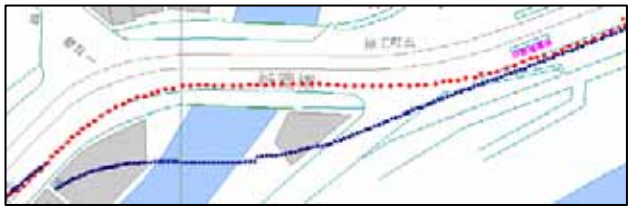

Fig. 6. Data fusion result for absolute road positioning

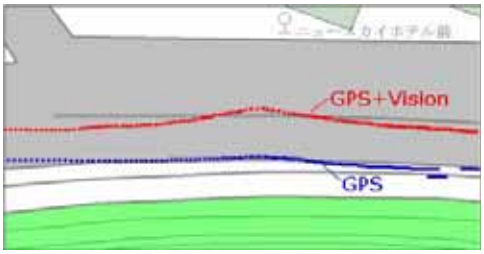

Fig. 7. GSP+Vision vs. GPS for road positioning

4.2 Estimation of Heading Angle and Offset

Because the measuring errors come from both GPS data (about 1.5m to 5m) and digital road map data itself (Zenrin® Z-Map Town II we used is based on 1:25000 digitized city map), calculating heading angle and offset to the road center line from image data is essential to the final synthesizing accuracy and visual performance when virtual objects are superimposing to the real road scene.

Fig.7 shows an example that even accurate GPS data has to be corrected by vision based estimation result. Image samples are shown in Fig.5(1) and Fig.5(2) while our test car changed to the right lane to avoid a stopped vehicle and returned back to the left lane later. GPS trace positions are plotted as "■" on the digital map as shown in Fig.7 and it is obvious incorrect. With our proposed camera pose estimation algorithm, test car's offset distance and heading angle to road center line were calculated and the new GPS+Vision positions are plotted as "●" in Fig.7.

4.3 On-Board Camera Pose Estimation Result

An example of lane changing is shown as the second scene of Fig.5. We changed to the right lane to avoid a stopped car and returned back to the left lane after. Fig.8 shows part of the angular rate estimation results of our hybrid approach. The results here has already been eliminated the original displacement between the camera and the vehicle reference frame. A very stable yaw rate change between frame #20 to frame #345 corresponds to the fact of changing lane as described above. Turbulence of roll and pitch angle rates between frame #720 to #760 can also be verified by the fact of uneven road surface near the intersection.

Fig.9 shows the result of integrated roll, pitch and yaw angle with respect to the world coordinate system. Fusion of vision result and gyro data makes our algorithm stable and robust, especially during the road intersection part and dense traffic scene when vision approach cannot work properly because road lane-markers were occupied or too complex. The accuracy of proposed algorithm can also be verified by the fact that the track of integrated Yaw angle perfectly matches with vehicle's absolute road position data obtained by GPS/gyro sensor (see Chapter 3.1).

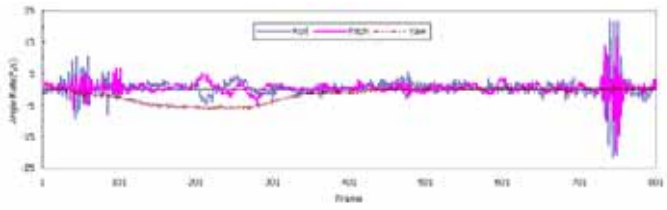

Fig. 8. Estimated result of rotation angle rate

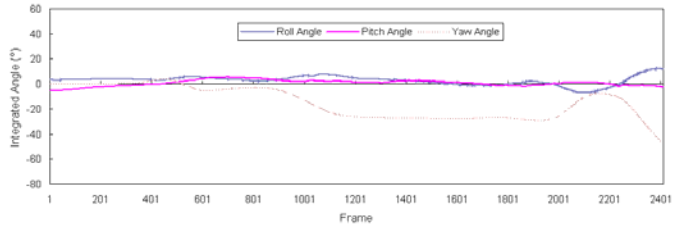

Fig. 9. Integrated pose angle result of our algorithm

4.4 Visual Perceived Performance Test

To generate virtual traffic indicators, bulletin boards and landmark icons, OpenGL® is employed here since it is not only a 3D modeling tool, but also a powerful 3D rendering engine to project virtual objects onto real-images' overlay with the proper camera pose data and all intrinsic and extrinsic parameters.

The following navigation information is extracted from the digital map: 1) road nodes location and segment attributes (name, level, lane info, etc.); 2) intersection location, names and crossing angles of the roads intersected; 3) landmarks, buildings and other value-added objects information (hospitals, gas stations, shopping centers, restaurants, etc.).

All information is dynamically extracted according to the current location (within certain range) and driver's preference. 3D objects are generated depending on its category: road information such as speed limits, direction indicators are modeled as virtual road paintings and are located on the road surface, road names and intersection information are modeled as virtual traffic bulletins mounted on a certain height above the road. can be recognized easily. Fig.10 shows some superimposing results by projecting the virtual objects on the real image overlay.

Since the ground truth of camera's extrinsic parameters is almost impossible to obtain, the following visual perceived performance tests were carried out to evaluate the accuracy of proposed algorithm.

Three typical road scenes were chosen for evaluation: an urban road with clearly marked lanes which is ideal for vision-based camera pose estimation, a widely open countryside road where GPS data was very accurate, and a one-lane downtown street with complex road markings and uneven road surfaces which was the most difficult but common scene for road navigation. We picked up some POIs (Point Of Interests) from each road and obtained their latitude and longitude data from digital road map. All POIs are visible along the driving route and no occlusion is considered in the perceived performance tests.

After estimating each frame's camera pose data, we converted the current road position and POIs' latitude/longitude data (which were based on Tokyo Datum) to the Euclidean planar coordinate system. With the calculated camera's extrinsic parameters, POIs' WCS coordinates were transformed to the camera based CCS coordinates and then projected to the image plane. An icon will be rendered on each POI's projection position. Icon's size and orientation is determined by POI's CCS coordinates. The movie files of evaluation results can be downloaded from the following web site: **http://navi.cs.kumamoto-u.ac.jp/~hu/ITS/image/.**

The following criterion was defined to evaluate the POI projection's accuracy:

$$Q = \{\left\| \vec{p}_{POI} - \vec{D}_{POI} \right\| < 20\} \quad (9)$$

where $\vec{p}_{POI}$ is the calculated POI's projection position and $\vec{D}_{POI}$ is manually extracted POI's position from each frame. We used a relative wide range to cover the slight variety of the reference position $\vec{D}_{POI}$ itself in each frame. If the distance is less than 20 pixels, we consider the estimation result as accurate.

All three scenes evaluation results are given in Table 1~3. Recall ratio is the percentage of accurately projected frames via total frames appeared. The results of the three scenes verify the high accuracy of camera pose estimation data, especially on the clearly marked straight roads. Displacements were relatively big on the complicated marked road scene like intersection left-turn arrow (Road Indicator C) in Scene #3 because vision sensor could not provide reliable road shape data due to the complicated intersection lane markers. It is the same reason for Clinic A and Clinic B in Scene #3. Discontinuity of GPS data was happened in Scene #1 due to a possible signal multi-reflection. POI position was updated by 3D gyro sensor only during this period. When GPS signal recovered, it caused a sudden movement of POI icon (Shop C in Scene #1). A simple weight-average filter should solve the problem.

4.4 Survey On the System Performance

Total of 50 testers are randomly selected with different age, sex, driving history and experience for the system performance test of VICNAS. The survey was carried out in our simulation environment where navigation movies recorded from both commercial navigation system and our VICNAS system were shown to the testers separately through a 19inch LCD monitor. Three deferent road scenes were chosen for the test and we selected the most recent off-the-shelf navigation system (Toyota DVD-Navi NDCN-D54, as shown in Fig.11) as our comparing target.

Fig. 12. shows the survey result on the item of *Understandability* of these two systems. Comparing with the current navigation system, VICNAS is obviously much easy to understand and the user-friendly interface gives it more potential advantages for future navigation.

The detail evaluation results on the different system features of VICNAS, like *Displaying Performance*, *Operating Safety* and *Convenience* are shown in Table 4, where point 5 means the best and point 1 means the worst. Very positive result can be seen from most of the system features, while some improvements are necessary in the displaying performance and system operating safety.

Table 1. An urban road with clearly marked lanes (Scene #1, total 2200 frames, 6 POIs)

POI	Longitude	Latitude	Elevation (m)	Recall Ratio (%)	Average Displacement (pixels)
Gas Station A	130”45’08.725	32”48’14.044	6.0	100	4.7
Shop A	130”45’20.042	32”48’08.982	6.0	82.3	11.2
Shop B	130”45’24.996	32”48’08.014	8.0	94.4	8.2
Shop C	130”45’27.183	32”48’07.136	12.0	100	4.9
Gas Station B	130”45’40.005	32”48’05.009	6.0	100	7.6
Restaurant A	130”45’42.106	32”48’04.436	8.0	100	9.1

Table 2. A widely open countryside road (Scene #2, total 600 frames, 1 POI)

POI	Longitude	Latitude	Elevation (m)	Recall Ratio (%)	Average Displacement (pixels)
Road Indicator A	130”48’34.876	32”50’13.812	2.0	100	7.6

Table 3. A one-lane downtown street with complex road markings and uneven road surfaces (Scene #3, total 1116 frames, 4 POIs)

POI	Longitude	Latitude	Elevation (m)	Recall Ratio (%)	Average Displacement (pixels)
Road Indicator B	130”41’53.042	32”47’22.806	2.0	100	2.1
Clinic A	130”41’50.804	32”47’23.187	6.0	76.0	17.2
Clinic B	130”41’50.028	32”47’23.036	6.0	77.7	14.5
Road Indicator C	130”41’43.648	32”47’20.634	2.0	63.8	23.5

Fig. 10. Superimposing results of virtual navigation information onto real images

Fig. 11. Output of recent commercial navigation system

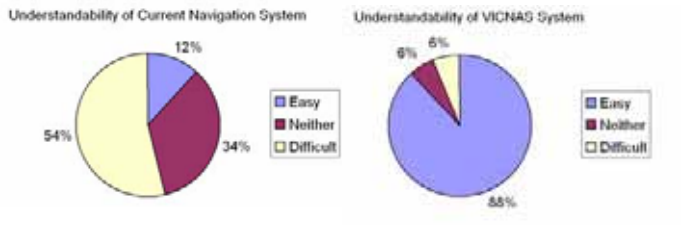

Fig. 12. Survey result on system understandability

Fig. 13. Output of refined VICNAS system

Table 4. Detailed survey result on different system features of VICNAS

Items \ Points	1	2	3	4	5	Total
Understandability	0	3	3	12	32	50
Displaying Performance	0	0	10	18	22	50
Operating Safety	0	2	15	19	14	50
Convenience	0	0	9	24	17	50

We have refined the output interface according to the survey result. 1) Navigation indicators and icons are made semi-transparent, and more colorful; 2) Animation of indicators and icons are erased to avoid disturbing drivers; 3) DTD (Distance To Destination) and DTT (Distance To the next Turn) indicators are added to properly navigate drivers. Fig.13. shows the refined system output.

5. Conclusion

This paper presents a novel framework of vision-based road navigation system, which superimposes virtual 3D navigation indicators and traffic signs onto the real road scene in an Augmented Reality (AR) space. To properly align the virtual object with real world, this paper proposed a hybrid camera pose tracking system that combines vision, GPS and 3D inertial gyroscope technologies. The fusion approach is based on our PMM (parameterized model matching) algorithm, in which the road shape model is derived from the digital map referring to GPS absolute road position, and matches with road features extracted from the real image. Inertial data estimates the initial possible motion, and also serves as relative tolerance to stable the pose output. The algorithms proposed in this paper are validated with the experimental results of real road tests under different conditions and types of road.

Error analysis will be one of the most important issue to solve before moving to the next stage. Since the ground truth of camera's extrinsic parameters is almost impossible to obtain, we plan to adopt You's method[10] to measure the difference of land marker's projection and the position automatically extracted from the image, under different focal length and distance.

Since the proposed algorithm currently is only considered to be worked in good weather condition, with paved road surface and relative accurate digital road map, the system will not work well or will be less accurate when drives in bad weather like heavy rain or snow, non-paved roads, or without road shape information like newly built roads and no-map area. All these options as well as on-line calculation will be considered in our future work.

There are still some special road shape segments that are not covered by our algorithm of road model matching, such as intersection and diversion, which are also essential for the on-road navigation. 3D road shape is also another interesting topic and it will become more commercially valuable when the 3D digital map data is available in the near future. Our interests will be continuously focused on these topics as well as the real-time computation and implementations in AR world.

6. Acknowledgments

This work was supported in part by the FS research grant from JST (Japan Science and Technology Agency).

7. References

[1] R.T. Azuma. "A survey of augmented reality," In *Presence: Teleoperators and Virtual Environments 6*, 4, pp.355-385 (1997).

[2] M. Bajura, H. Fuchs, and Ohbuchi. "Merging virtual objects with the real world: Seeing ultrasound imagery within the patient," *Computer Graphics*, pp.203-210, July 1992.

[3] Z. Hu, K. Uchimura, "Solution of Camera Registration Problem via 3D-2D Parameterized Model Matching for On-Road Navigation", *International Journal of Image and Graphics*, Vol. 4, No. 1, pp.1-18 (2004)

[4] E. Foxlin, "Inertial Head-Tracker Sensor Fusion by a Complementary Separate-Bias Kalman Filter", *Proc. of IEEE Virtual Reality Annual International Symposium*, pp.184-194, 1996.

[5] E. Foxlin, M.Harrington, and G. Pfeifer, "Constellation: A Wide-Range Wireless Montion-Tracking System for Augmented Reality and Virtual Set Applications"*, Proc. Of GRAPHICS 98*, 1998.

[6] B. Horn and E. Weldon. "Direct Mothods for recovering motion," Intl. Journal of Computer Vision, Vol.2, pp. 51-76 (1988)

[7] X. Zhuang, R. Haralick, and T. S. Huang, "Two-view motion analysis: A unified algorithm," *Opt. Soc. Am.*, 3(9), pp. 1492-1500, 1986.

[8]R.M. Haralick, H. Joo, R. C. Lee, X. Zhuang, V. G. Vaidya, and M. B. Kim. "Pose Estimation from Corresponding Point Data," *IEEE Transactions on Systems, Man and Cybernetics*, 19(6), pp.1426-1446, November-December 1989.

[9] L. Chai, K. Nguyen, B. Hoff, and T. Vincent, "An Adaptive Estimator for Registration in Augmented Reality", *Prof. of IEEE International Workshop on Augmented Reality*, pp. 23-32, 1999.

[10] S. You and U. Neumann. "Fusion of vision and gyro tracking for robust augmented reality registration", *Proc. of IEEE Conference on Virtual Reality*, pages 71-78, Japan, 2001

[11] A. Valinetti, A. Fusiello, V. Murino. "Model tracking for video-based virtual reality," *Proc. 11th Intl. Conf. On Image Analysis and Processing*, pp.372 –377 (2001)

[12] R. Gregor, M. Lutzeler, M. Pellkofer, K.-H. Siedersberger, and E. D. Dickmanns. "EMS-Vision: A Perceptual System for Autonomous Vehicles," *IEEE Trans. on Intelligent Transportation Systems*. 3(1), pp. 48-59, March 2002.

[13] R. Hooke and T. Jeeves. "Direct search solution of numerical and statistical problems," *Journal of the Association for Computing Machinery (ACM)*, pp.212-229 (1961).

[14] Z. Hu, K. Uchimura, "On-board Camera Pose Estimation in Augmented Reality for Direct Visual Navigation", *IS&T/SPIE's Electronic Imaging 2003*, Santa Clara, California, USA, pp.508-518 (2003)

[15] A. Kelly, "Essential Kinematics for Autonomous Vehicles", Technical Report of Carnegie Mellon University, CMU-RI-TR-94-14 (1994)

[16] T. Oike, "A White Road Line Recognition System using the Model-Based method", Technical Report of IEICE, PRMU99-221, pp.53-60 (2000-01)

[17] Z. Hu, K.Uchimura, "Recognition of Horizontal Shape Models for General Roads", Trans. of IEICE, Vol.J81-A, No.4, pp.590-598 (1998)

Zhencheng Hu received his B.Eng degree from Shanghai Jiao Tong University, China in 1992, and his M.Eng. degree from Kumamoto University, Japan, in 1998. He received his Ph.D. degree in System Science from Kumamoto University, Japan, in 2001. Dr. Hu has held various positions in computer science and machine vision industry. He is currently an associate professor with the Department of Computer Science, Kumamoto University, Japan. His research interests include camera motion analysis, augmented reality, machine vision applications in industry and ITS. Dr. Hu is a member of IEEE, and the Institute of Electronics and Information Communication Engineers of Japan (IEICE).

Keiichi Uchimura received the B. Eng. and M. Eng. degrees from Kumamoto University, Kumamoto, Japan, in 1975 and 1977, respectively, and the Ph. D. degree from Tohoku University, Miyagi, Japan, in 1987. He is currently a Professor with the Department of Computer Science, Kumamoto University. He is engaged in research on intelligent transportation systems, and computer vision. From 1992 to 1993, he was a Visiting Researcher at McMaster University, Hamilton, ON, Canada. Dr. Uchimura is a Member of the Institute of Electronics and Information Communication Engineers of Japan.

- *Received date: April 19 2005*
- *Received in revised forms : July 23 2005, October 17 2005, January 26 2006*
- *Accepted date: January 26 2006*

- *Editor: Katsushi Ikeuchi*